DOCUMENT RESUME

ED 277 282 FL 016 390

AUTHOR Read, John; Nation, Paul
TITLE Some Issues in the Testing of Vocabulary Knowledge.
PUB DATE May 86
NOTE 28p.; Paper presented at a language testing symposium in honor of John B. Carroll and Robert Lado (Quiryat Anavim, Israel, May 11-13, 1986).
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Achievement Tests; Check Lists; Classification; Comparative Analysis; Diagnostic Tests; *Evaluation Criteria; Expressive Language; Item Sampling; *Language Tests; Language Usage; Receptive Language; Second Language Instruction; *Test Construction; *Test Format; *Test Items; *Vocabulary; Vocabulary Development; Word Lists

ABSTRACT

A review of the literature on a variety of issues related to testing vocabulary knowledge in a second language addresses these topics: problems in estimating vocabulary size, including the related questions of what constitutes a word, how a sample should be selected, and what are the criteria for knowing a word; sampling the basic and specialized vocabulary of second language learners; diagnostic use of tests of vocabulary knowledge and the construction of such tests; the checklist as an alternative diagnostic measure; problems related to testing words in isolation; and the distinction between receptive and productive vocabulary knowledge. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED277282

SOME ISSUES IN THE TESTING OF VOCABULARY KNOWLEDGE

John Read and Paul Nation

English Language Institute
Victoria University of Wellington
Wellington
New Zealand

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

John Read

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at LT + 25: A Language Testing Symposium in Honor of John B Carroll and Robert Lado, Quiryat Anavim, Israel, 11-13 May 1986

FL016390

BEST COPY AVAILABLE

SOME ISSUES IN THE TESTING OF VOCABULARY KNOWLEDGE

John Read and Paul Nation
Victoria University of Wellington

Vocabulary is a component of language proficiency that has received comparatively little attention in language testing since the general move towards more integrative formats. The testing of word knowledge was a core element of the discrete-point philosophy but with changing ideas about the concept of language test validity it has tended to be neglected in favour of higher level skills and processes, so that vocabulary is seen as just one of the numerous elements that contribute to the learner's overall performance in the second language. However there seems to be a growing recognition of the importance of vocabulary and the need for more systematic vocabulary development for second language learners, many of whom are severely hampered in reading comprehension and other skills by a simple lack of word knowledge. Since the standard types of integrative test do not provide a direct assessment of this knowledge in a form that is useful for diagnostic purposes, we believe that there is a need to develop tests to determine whether specific learners have achieved a mastery of vocabulary that is sufficient for their needs and, if they have not, what can be done pedagogically to help them.

Our interests in vocabulary testing have both a practical and a more theoretical focus. On the practical side, the English Language Institute in Wellington - where we work - has traditionally placed great emphasis on the acquisition of vocabulary in its English proficiency course for foreign students from Asia and elsewhere who are preparing to study in New Zealand universities. This emphasis derives in part from the results of studies by Barnard (1963) in India and Quinn (1968) in Indonesia, which

both provided evidence of the low level of English vocabulary knowledge among university students in Asia, even after extensive study of English at the secondary level. Quinn found, for example, that the average university entrant in his sample had a vocabulary of 1000 words after six years of study, which represented a learning rate of little more than one word for each class hour of English instruction. Such limited vocabularies were clearly inadequate to meet the demands of English-medium university studies. The Institute has thus given a high priority to intensive vocabulary learning in its proficiency course and has developed a variety of teaching resources for this purpose, including in particular the commercially published workbooks by Barnard (1971-75), who pioneered the work in this area. In this context, there is a particular need for diagnostic testing to assess the vocabulary knowledge of specific learners in order to assist in making placement decisions and in designing effective programmes of vocabulary development for the various groups on the course. On a more theoretical level, we are investigating the effectiveness of various types of vocabulary test as tools in ongoing and planned research studies on vocabulary size, the nature of vocabulary knowledge and the role of vocabulary in reading comprehension.

Problems in Estimating Vocabulary Size

The basic question in a diagnosis of a learner's vocabulary knowledge is simply: how many words does the learner know? The question is easy to frame but rather more difficult to answer. Since comparatively little work has been done in this area with second language learners, we need to turn to the literature on the vocabularies of native speakers in order to clarify the issues

involved.

There has been a great deal of research on vocabulary size, extending back to the end of the last century, and learned speculation on the subject goes back much further. However, as Anderson and Freebody (1981) point out in a review of the research, it is difficult to have much confidence in the results of all these studies because they have yielded such widely varying estimates of the number of words known by specified groups of subjects, such as children of a particular age, college students or educated adults. Even for quite young children, say five-year-olds, we can find estimates ranging all the way from 2500 to 26000 words (quoted in Lorge and Chall, 1963). In the case of university students, the discrepancy is correspondingly large, with the low estimate being Seashore's (1933) figure of 19000 as compared with Diller's (1978) all-time high total of 216000 words. Such discrepancies make it clear that there have been significant methodological problems in this type of research that have either not been recognized at all or have been treated in various ways by different researchers.

The methodological issues have been discussed in detail elsewhere (see, e.g., Lorge and Chall, 1963; Anderson and Freebody, 1981) and need only be summarized here.

(a) What is a word?

The first problem is simply to define what a word is. For instance, are depend, depends, depended and depending to be classified as one word or four? And how about dependent, dependant, dependence and dependency? That is, one has to decide whether a 'word' is an individual word form or a word family (or lemma) consisting of a base form together with the inflected and derived forms

that share the same meaning. Similarly, the status of such items as proper nouns, compound words, abbreviations, obsolete words and slang expressions needs to be considered. Including all such forms as separate words will clearly increase the estimate of vocabulary size, whereas a more conservative approach results in a substantially lower figure. The latter approach would seem to be much more realistic, but it requires a careful definition of criteria for grouping words into families and even then there are difficult problems of classification to deal with. A useful discussion of this issue can be found in Nagy and Anderson (1984).

(b) How should a sample of words be selected?

Since it is doubtful whether all the words in the language can be listed, let alone tested, it is necessary to find some basis for selecting a representative sample of words to be used in making the estimate of vocabulary size. Mostly commonly, a dictionary has been used for this purpose, with the result that the size of the estimate has a predictable relationship to the size of the dictionary. When a larger dictionary is used, the subjects are credited with knowing a greater total number of words, even if the actual number of words tested remains constant. This reflects the fact that even the largest dictionaries in existence cannot claim to contain all the words in the language - in principle, such comprehensiveness is impossible to achieve - and so there is no absolute basis for making the estimates. A further problem is that a dictionary is not a very satisfactory sampling frame. As Lorge and Chall (1963) and others have shown, systematic sampling at fixed page intervals throughout a dictionary produces a sample in which very frequent words are overrepresented, because these words typically have both multiple listings and much

longer entries than low-frequency words and thus they are more likely to be selected. This in turn contributes to an inflated estimate of vocabulary size, since frequent words will be better known to the subjects than infrequent words.

(c) What is the criterion for knowing a word?

Once a sample of words has been selected, it is necessary to determine whether each word is known or not by means of some kind of test. In practice, the criterion for knowing the word has been quite liberal, since the researcher has had to survey a large number of words in the time available for testing. Thus test formats such as checking, multiple choice and matching have been the most commonly used (cf. Sims, 1929). This raises the question of whether simply ticking a list or making a correct response to a single multiple choice item is a valid basis for crediting someone with knowing a word. Being able to associate a word and a definition is only one aspect of vocabulary knowledge. We need to take account of the fact that words can have multiple meanings and, conversely, a person's knowledge of a word may be partial rather than complete. In an analysis of the components of word knowledge, Cronbach (1942) identified five sorts of behaviour involved in understanding a word. These were generalization (being able to define the word); application (selecting an appropriate use of the word); breadth of meaning (recalling the different meanings of the word); precision of meaning (applying the word correctly to all possible situations); and availability (being able to use the word). In a more recent article, Richards (1976) discusses several other aspects of knowing a word, such as its relative frequency, its syntactic properties, its connotations and its links with other words in semantic networks. Such analyses make it clear that

the typical estimate of vocabulary size is based on a crude measure of vocabulary knowledge, and that these broad surveys should be complemented by more in-depth studies of how smaller sets of key vocabulary items are known.

The three methodological issues outlined above are all relevant to testing the vocabulary knowledge of second language learners, but only the second and third ones will be discussed in subsequent sections of the paper. With regard to the first question, we will simply take it that the lexicon can be classified into lemmas (or word families), which can be represented for testing purposes by a base word. Thus, we assume that, if one knows the base word, little if any additional learning is required in order to understand its various inflectional and derived forms. Obviously this assumption is not always justified: derived forms may be substantially different in meaning from the base word; in which case there may be different lemmas involved.

Sampling the Vocabulary of Second Language Learners

One of the problems in estimating the vocabulary size of native speakers is that from quite a young age they know such a large a diverse range of words. This is why researchers have generally preferred to sample from a comprehensive dictionary, as noted above. An alternative approach which involves sampling words progressively from the higher to lower levels of a word frequency count has not proved very successful, because of the limited coverage of the existing frequency counts and the fact that native speakers know many words that do not find their way into such counts. However, this latter approach is more appropriate with second language learners, who have much less exposure to the language and whose communicative

needs are also typically more limited. For many groups of learners, especially those in EFL countries with little or no exposure to the language outside the classroom, the General Service List (West, 1953) represents a fairly complete sampling frame of the words they are likely to know, even after several years of study. And in fact a number of ESL vocabulary studies (e.g. Barnard, 1963; Quinn, 1968; Harlech-Jones, 1983) have used the list for this purpose.

Beyond the minimum vocabulary of the General Service List, it is necessary to take account of the needs and interests of specific groups of learners in planning for vocabulary teaching and testing. In fact it is also necessary to give a justification for focusing on any particular list of vocabulary items above the basic level at all, since it is well known that the vocabulary knowledge of individuals (both native and non-native speakers) varies considerably according to their personal interests and experiences, intelligence, linguistic and cultural background, education and so on. Our specific interest is in the vocabulary of English for academic study at the tertiary level. The task here is to identify and teach the set of words (often referred to as 'subtechnical' vocabulary) that occur frequently in academic discourse across various disciplines. Knowledge of the meanings of these words is normally assumed by lecturers and authors in a particular academic field and among other things this vocabulary has a crucial role in defining the technical terms of each field of study.

A number of specialized word lists for academic English have been produced. Typically these are based on a count of words occurring in university textbooks and other academic writing material, taking into account the range of disciplines in which the words

are found as well as the number of occurrences. The list is compiled by excluding both high frequency general words (such as those in the General Service List) and low frequency, narrow range words which consist largely of technical terminology. The most comprehensive work along these lines is that of Barnard, who has not only prepared two 1000-word lists (in Nation, 1984b) but also written a series of widely used workbooks (Barnard, 1971-1975) to help students to learn the words. Other similar, though shorter lists were compiled by Campion and Elley (1971) and Praninskas (1972). Two more scholars, Lynn (1973) and Ghadessy (1979) adopted a different approach, by scanning student copies of textbooks to identify and count words that were frequently annotated with a mother-tongue translation or some other explanation. These words turned out to be very much the same ones that were included in the other lists. As Lynn (1973:26) noted specifically, it was the general academic words, rather than technical terms, that appeared to be the most difficult for the students.

Xue and Nation (1984) have combined the lists of Campion and Elley, Praninskas, Lynn and Ghadessy into a single University Word List, which shares much in common with the Barnard lists but has the advantage of being derived from a broader range of frequency counts. The list is accompanied by sublists which classify the words according to frequency and semantic criteria and also include common derivatives of the base forms.

For our purposes the academic word lists - together with the General Service List - form an inventory of high frequency words that commonly occur in academic English and account for a high proportion of the words in any particular academic text. These are the words that require individual attention

from teachers and learners in an EAP proficiency course. The lists also constitute a satisfactory sampling frame for diagnostic testing aimed at evaluating the adequacy of the leanrer's vocabulary knowledge for undertaking academic study. The high frequency words need to be distinguished from low frequency words, which are also encountered in academic study but which are too numerous and specialized to be formally taught by the language teacher. What learners need to deal with unknown low frequency words are strategies such as guessing meaning from contextual clues, using knowledge of prefixes, roots and suffixes or simply ignoring the word if appropriate. Thus with high frequency words it is necessary to test knowledge of the words individually, whereas in the case of low frequency words it is the skills in coping with such words in general that need to be assessed.

A Diagnostic Test of Vocabulary Knowledge

A first attempt to undertake diagnostic testing of vocabulary knowledge along the lines outlined above is represented by the Vocabulary Levels Test (Nation, 1983). This instrument was designed to assess knowledge of both general and academic vocabulary; therefore it includes samples of words at five frequency levels: 2000 words, 3000 words, 5000 words, the university word level (above 5000 words) and 10000 words. Words were selected on the basis of the frequency data in Thorndike and Lorge, with cross-checking against the General Service List (for the 2000-word level) and Kucera and Francis (1967). The one exception was the university word level, for which the specialized count of Campion and Elley (1971) was used. (This list excluded the first 5000 words

of Thorndike and Lorge.) The test employs a word-definition matching format although, in a reversal of the standard practice, the testees are required to match the words to the definitions. That is, the definitions are the test items rather than the words. At each of the five levels, there are 36 words and 18 definitions, in groups of six and three respectively, as in the example below:

1	apply		
2	elect	_____	choose by voting
3	jump	_____	become like water
4	manufacture	_____	make
5	melt		
6	threaten		

This slightly unconventional format was developed with the aim of having an efficient testing procedure that involved as little reading as possible and minimized the chances of guessing correctly. It was considered that, although there were only 18 words for each level, in fact 36 words would be tested because the testees' natural test-taking strategy would be to check each word against the definitions given in order to make the correct matches. This was only partly confirmed by observation of individual testees as they took the test during the tryout phase. The testees did adopt that strategy but only in sections of the test that they found difficult; with easy items they focused directly on the correct words and largely ignored the distractors.

All the words in each group are the same part of speech, in order to avoid giving any clue as to meaning based on form. On the other hand, apart from the correct matches, care was taken not to group together words and definitions that were related in meaning. The test is designed as a broad measure of word knowledge and it was not intended to require the testees to differentiate

between semantically related words or to show an awareness of shades of meaning.

The test has proved to be a very useful tool for diagnostic purposes. We have basic statistical data available from the administration of the test during our three-month ELI English Proficiency Course in the summer of 1984-85 (see Table 1). The test was given at the beginning of the course, to assist with placement and course planning decisions, and again at the end, in order to look at the stability of the instrument and the possibility that it would reflect the effects of instruction. For both administrations, the reliability coefficients were very satisfactory (0.94 and 0.91 respectively) and there was a clear pattern of declining scores across frequency levels from highest to highest. However the means for the 5000-word and university levels were very close at the beginning of the course and in fact their order was reversed in the second administration, for reasons that will be discussed in a moment.

In order to provide more systematic evidence of the validity of the division by levels, a Guttman Scalogram analysis (Hatch and Farhady, 1982) was undertaken on the two sets of scores. A score of 16 was taken as the criterion for mastery of the vocabulary at a particular level. The scaling statistics are given in Table 1. They show that in both cases the scores were highly scalable. That is to say, a testee who achieved the criterion score at a lower frequency level - say, the 5000-word level - could normally be assumed to have mastered the vocabulary of higher frequency levels - 2000 and 3000 words - as well.

Table 1: Results of the Vocabulary Levels Test

(1984-85 ELI Proficiency Course)

(N = 81)

1st Administration (Beginning of Course)

Level	2000	3000	5000	University	10000	Total
No. of items	18	18	18	18	18	90
Mean	16.4	15.7	12.3	11.6	6.7	62.8
S.D.	2.3	3.3	4.3	4.7	3.5	15.3

Guttman Scaling:

Order of Levels	Statistics
2000	C_{rep} = 0.93
3000	MM_{rep}= 0.37
5000	Scalability = 0.90
University	
10000	

2nd Administration (End of Course)

Level	2000	3000	5000	University	10000	Total
No. of items	18	18	18	18	18	90
Mean	17.0	16.6	13.9	14.1	8.8	70.3
S.D.	1.1	2.3	3.6	3.8	3.6	11.7

Guttman Scaling:

Order of Levels	Statistics
2000	C_{rep} = 0.92
3000	MM_{rep}= 0.48
University	Scalability = 0.84
5000	
10000	

What is not so satisfactory from a theoretical standpoint is that there are two different scales here, because the 5000-word and university-word levels reverse their order from one administration to the other. There are various possible explanations for this. First of all, the university level is based on the specialized frequency count of Campion and Elley, whereas the other four levels are all derived from the more general Thorndike and Lorge count. This places the university level somewhat outside the sequence formed by the other levels.

Secondly, the students taking the proficiency course are quite heterogeneous and the overall results mask some significant differences among subgroups in the population. For instance, one group consists of teachers of English from EFL countries in Asia and the South Pacific preparing for a Diploma course in TESL during the following academic year. These teachers tend to score higher at the 5000-word level than the university level, reflecting their familiarity with general English, including literary works, and their relative lack of familiarity with academic or technical registers. This was particularly evident at the beginning of the course but was less noticeable at the end, presumably because of their exposure during the course to academic writing and their study of the University Word List.

On the other hand, another identifiable subgroup comprises a small number of Latin American students who are native speakers of Spanish or Portuguese. Unlike most of the English teachers, they are university graduates coming to New Zealand for postgraduate studies in engineering or agriculture. The students almost all had substantially higher scores at the university-word level than the 5000-word level. One way to explain this is in terms

of their academic background, even though it was not in the medium of English. Another factor is that a high proportion of the words at the university level are derived from Latin and are therefore likely to have cognate forms in Spanish and Portuguese, with the result that Latin American students would have a certain familiarity with them without having learned them as English words.

A third reason for the shift in the order of levels from the first to the second administration is that a great deal of attention is paid to academic vocabulary during the proficiency course. Almost all of the groups work through two of the workbooks in the Advanced English Vocabulary series (Barnard, 1971-75) and the University Word List is also used as the basis for vocabulary learning activities. Thus, if the test is sensitive to the effects of instruction during the course, one might expect a relatively greater improvement in scores at the university word level than at the 5000 word level. There is some evidence from the pretest and posttest means that this was the case.

An Alternative Measure: The Checklist

Although the Vocabulary Levels Test has proved to be a diagnostic tool, we are aware of at least three possible shortcomings.

(a) It tests a very small sample of words at each level, even if we accept that 36 words are tested rather than just 18.

(b) The matching format requires the testees to match the words with dictionary-type definitions, which are sometimes awkwardly expressed as the result of being written within a controlled vocabulary. Learners may not make sense of words in quite the analytic fashion that a lexicographer does.

(c) While the format was modified to reduce the role of memory and test-taking strategy, there is still a question of the influence of the format on testee performance.

Thus, as an alternative format, we have been looking at what might be regarded as the simplest and purest vocabulary test of all. This is the checklist (also called the yes/no method), which simply involves presenting learners with a list of words and asking them to check (tick) each word that they "know". The exact nature of the task depends - more so than with most other tests - on the testees' understanding of what they are being asked to do, and therefore both the purpose of the test and the criterion to be used in judging whether a word is known need to be carefully explained. As with any type of self-evaluation, it is not suitable for grading or assessment purposes, but it has definite appeal as an instrument in vocabulary research, especially since it is an economical way of surveying knowledge of a large number of words.

A review of the literature on the checklist method reveals that it is one of the oldest approaches to the testing of vocabulary: Melka Teichroew (1982:7) traces it back as far as 1890. There have been two main concerns in the research: (1) how valid the results of a checklist test are; and (2) how to control for a presumed tendency among students to overstate their knowledge, by ticking words that they do not actually know.

Earlier studies tended to concentrate on the first question. For example, Sims (1929) compared a checklist test with three other tests - multiple choice, matching and identification (oral interview) - of the same words and found that it did not correlate very well. Since the three other tests were highly intercorrelated,

he concluded that the checking test was not measuring knowledge of word meaning but simply familiarity with the words from having frequently encountered them in reading and school work. It should be noted that his subjects were school children in grades 5 to 8 who were perhaps not as able as older learners to distinguish familiarity with the form of the word from knowledge of its meaning.

On the other hand, Tilley (1936) found evidence for the validity of the checklist in his study of the relative difficulty of words in three standardized tests of vocabulary for children at three grade levels of the elementary school. He calculated difficulty scores for each word from both the checklist (or "self-appraisal test", as he called it) and a conventional multiple-choice test. High correlations were found between the two sets of scores and Tilley interpreted this as evidence for the concurrent validity of the self-appraisal method. When separate analyses were performed for various subgroups in the sample, it was shown that there was a somewhat stronger relationship between the two measures in the higher grades and at higher levels of intelligence.

However the implication that the checklist was a superior method with older and more intelligent subjects was not supported in a study by Cronbach (1942). He prepared a list of 60 technical terms in psychology and presented it to two university psychology classes. The students were told to check all those words whose meaning they understood in the context of psychology. As a validation measure, the students were asked the following day to write a 20-word explanation of some of the words from the list. The results showed that the checklist responses gave a poor indication of how well the students understood the terms. In addition, the checklist was only a rough guide to the relative

difficulty of the terms for these students. It should be noted that the subjects in this study apparently had no previous background in psychology (though Cronbach is not explicit on this point) and many of their incorrect responses were explanations of the general meaning of the terms rather than their technical definitions in psychology.

Two more recent studies have produced more favourable results for the checklist method. As part of their project to develop an academic vocabulary list based on a word frequency count of university textbooks, Campion and Elley (1971) asked senior high school students to rate each word according to whether they could attribute a meaning to it if they encountered it in their reading. The percentage of positive responses was taken as an index of the familiarity of the word for university-bound students. The ratings correlated reasonably well (at 0.77) with the results of a word-definition matching test.

One innovation in Campion and Elley's study addressed the second issue identified above: how to control for a persistent tendency to tick words that were not actually known. The subjects of this study were divided into groups, which each rated a different subset of the words in the list. However each sublist included a number of "anchor" words, which were thus rated by all of the subjects. The mean ratings of the anchor words were calculated and these were used in a norm-referenced fashion to evaluate the performance of the various groups. When a group was found to have rated the anchor words significantly differently from the overall means, the ratings of the words on its sublist were adjusted as appropriate.

Another approach to this issue has been adopted by Anderson

and Freebody (1983). Their solution to the problem was foreshadowed in the previously cited study by Cronbach (1942), who included in his checklist of psychological terms a nonsense word: denomia (and, sure enough, two students ticked it). Anderson and Freebody developed this idea by preparing a vocabulary checklist containing a high proportion (about 40 per cent) of "nonwords". These were created either by changing letters in real words (e.g. porfame from perfume) or by forming novel base-and-affix combinations (e.g. observement). The ticking of these nonwords was taken as evidence of a tendency to overrate one's knowledge of the real words, and a simple correction formula was applied (similar to a correction for guessing) to adjust the scores accordingly. The corrected scores were found to correlate much more highly with the criterion - the results of an interview procedure - than did scores on a multiple choice test of the words. In a subsequent study on learning words from context, Nagy, Herman and Anderson (1985) used a similar checklist test as a measure of the subject's prior knowledge of the target words. In this case complete nonwords such as felinder and werpet were included in the checklist, in addition to the other two types, and it was only this third category of nonwords that was used in making the corrections of subjects' scores.

The checklist approach does not appear to have been applied to any significant extent to studies of the vocabulary of second language learners. Melka Teichroew (1982) mentions in passing one Dutch study but generally casts doubt on the validity of the procedure. However, in the broader context of second language testing, it fits in well with recent work on self-assessment by adult foreign language learners (Oskarsson, 1980; von Elek,

1982). In fact, the vocabulary section of von Elek's (1982:22) test of Swedish as a second language includes very similar tasks, except that the words are contextualized in short sentences rather than being presented in isolation.

In summary, then, the checklist test has much to recommend it as a broad measure of vocabulary knowledge, especially if it incorporates a correction procedure for overrating. The simplicity of the test is a significant advantage. As Anderson and Freebody note, "it strips away irrelevant task demands that may make it difficult for young readers and poor readers to show what they know" (1983:235). (Their subjects were in fact fifth grade students.) It does not require the kind of testwiseness that influences performance in multiple choice or matching tests. A related attraction is that a much larger number of words can be assessed by the checklist in a given period of time, as compared to other types of vocabulary test.

Thus, we plan to develop a checklist version of the levels test in order to investigate its suitability as a diagnostic measure in our own work. In fact an informal type of checklist was used during the development of the matching test as an indication of the validity of the test, with encouraging results. Having investigated the literature on the checklist more thoroughly, we are optimistic that it will prove to be a valid and practical instrument.

Testing Words in Isolation

A question that arises in the use of both matching and the checklist is whether it is justifiable to test the words in isolation, since it tends to be taken for granted these days that words

should always be tested in context. There appear to us to be two main justifications for testing in isolation. The first is essentially practical: it is an efficient way of testing words. The advantages of the checklist test, as summarized above, derive substantially from the fact that the words are isolated. Unnecessary reading is eliminated and the testees' attention is focused directly on the task in hand. In addition, a larger number of words can be covered in a given period of time.

The second justification comes from a review of the literature on foreign language vocabulary learning and in particular the question of whether words are best studied in lists or in context (Nation, 1984:135-137). Experiments that have compared initial learning of words in context with learning word pairs (foreign word-English translation) have not produced results favouring learning in context. Admittedly there are methodological problems with these studies, especially concerning the nature and function of the context that is provided in the learning-in-context condition. However, despite the appeal of the idea that vocabulary should always be taught and tested in context, there is at present a lack of empirical evidence for it in the psycholinguistic literature. Learning words from uncontextualized lists can be a highly effective method of vocabulary acquisition, at least in the initial stages.

Our main concern is not to push the case for testing in isolation but rather to argue against the assumption that such tests are no longer worthy of serious consideration. We believe that this kind of test does have a role in the testing of vocabulary knowledge. The type of knowledge tested is one aspect of knowing vocabulary, but of course other aspects need to be assessed as

well.

Receptive and Productive Knowledge

In seeking ways to move beyond the rather limited measures of vocabulary knowledge discussed so far, we have been looking at the traditional distinction between receptive and productive vocabulary. Unfortunately, as Melka Teichroew (1982) points out, a survey of vocabulary studies that invoke this distinction reveals considerable confusion about its conceptual basis and how it should be operationalized. This is indicated at a basic level by the variety of terms used to refer to these concepts: receptive-productive; active-passive; comprehension-production; understanding-speaking; recognitional vocabulary-actual or possible vocabulary use. At the operational level Melka Teichroew shows that there is no consensus among researchers as to how the two types of vocabulary should be measured. In fact certain types of test, such as the checklist, multiple choice and translation, have been used by different researchers to measure both receptive and productive vocabulary. It comes as no surprise, then, that there is wide variation in their estimates of the relative size of the two types of vocabulary.

If we accept that all productive vocabulary is also known receptively, the real problem with the distinction is to decide what constitutes evidence that a word is part of a learner's productive competence. The strongest evidence presumably is that the word occurs in the learner's speech and writing, either spontaneously or in response to an elicitation device, such as a sequence of pictures or a topic nominated by the researcher. However a word count derived in this way will obviously be an underestimate of productive

vocabulary, since many words that could be used will not actually occur in the samples recorded by the researcher. Thus somewhat weaker evidence is often admitted. The learner is presented with the word and asked to use it appropriately in a sentence, give a definition or translate the word from L1 into L2. But less direct tests such as these can equally well be regarded as merely measuring receptive knowledge of the words. At best these tests indicate whether people are sufficiently knowledgeable about the words to be able to use them; they do not establish whether the words are actually used or not.

Our vocabulary levels test and the checklist test clearly focus on receptive (or passive) vocabulary knowledge. There are at least two reasons why we should feel some uncertainty about the adequacy of such a focus.

1 Corson (1985) argues that unless vocabulary is used actively in speech or writing then it is unlikely that learners will develop the cognitive framework which will allow effective use of this vocabulary in any of the four skills. Thus measuring receptive knowledge of the important university vocabulary may still not indicate whether learners have sufficient mastery of such vocabulary. A high score on a receptive test may mean that a teacher has to arrange further productive practice with that vocabulary. And in fact a recent study by McKeown et al (1985) indicates that for first language learners a high score on a word form-definition matching test does not provide reliable evidence that the learner can readily access the word and can understand sentences depending on knowledge of the word.

2 The receptive-productive distinction is of doubtful value.

The distinction is based on use: does (or can?) the learner use the word? This raises some interesting questions. If a learner uses a word, but uses it incorrectly in some way, is the word a part of the learner's productive vocabulary? If a learner knows a word very well and could use it but has never used it, is it a part of the learner's productive vocabulary? These questions arise because as teachers we are interested in knowledge as much as use. Rather than ask, "Is this word destined to be a part of the learner's productive or receptive vocabulary?" we should ask, "What features of this word need to be learned in order for the learner to know this word well?" The answer may well surprise us. Analyses of what it means to know a word come up with a list of aspects which include the word's sound, spelling, grammatical patterning, collocations, appropriacy, frequency, associations and meaning (Richards, 1976). Once a learner has a reasonable degree of familiarity with English, many of these factors involve little learning because they are predictable on the basis of the way previously learned words in English or the mother tongue behave. For example, if I tell you that <u>savoy</u> is "kind of cabbage with wrinkled leaves" you should be able to predict whether it is countable or uncountable, how to spell it (if you have not seen the word), what adjectives could go with it, and what other known words share its semantic field. It is also highly likely that you could immediately use this word if the opportunity arose.

Thus, research is needed to determine whether there is a continuum of knowledge for most words with certain aspects of knowledge in a fixed order on that continuum, which aspects of knowledge

should be tested to provide the most useful measure of vocabulary knowledge, and which test items do this most efficiently.

In spite of almost a century of research, our knowledge of vocabulary size and procedures for investigating vocabulary knowledge is still scanty. What knowledge we have comes mainly from studies of first language learning. There is clearly a need for further research in this area.

References

Anderson, R. C. and Freebody, P. 1981: Vocabulary knowledge. In Guthrie, J. T., editor, Comprehension and teaching: research reviews. Newark, Delaware: International Reading Association

_____ 1983: Reading comprehension and the assessment and acquisition of word knowledge. In Hutson, B., editor, Advances in reading/language research. Greenwich, Connecticut: JAI Press

Barnard, H. 1961: A test of P.U.C. students' vocabulary in Chotanagpur. Bulletin of the Central Institute of English 1, 90-100

_____ 1971-75: Advanced English vocabulary. Workbooks 1-3. Rowley, Massachusetts: Newbury House

Campion, M. E. and Elley, W. B. 1971: An academic vocabulary list. Wellington: New Zealand Council for Educational Research

Carroll, J. B., Davies, P. and Richman, B. 1971: The American Heritage word frequency book. New York: American Heritage

Corson, D. J. 1985: The lexical bar. Oxford: Pergamon

Cronbach, L. J. 1942: An analysis of techniques for diagnostic vocabulary testing. Journal of Educational Research 36, 206-217

Diller, K. C. 1978: The language teaching controversy. Rowley, Massachusetts: Newbury House

Ghadessy, M. 1979: Frequency counts, word lists, and materials preparation: a new approach. English Teaching Forum 17, 24-27

Harlech-Jones, B. 1983: ESL proficiency and a word frequency count. English Language Teaching Journal 37, 62-70

Hatch, E. and Farhady, H. 1982: Research design and statistics for applied linguistics. Rowley, Massachusetts: Newbury House

Kucera, H. and Francis, W. M. 1967: A computational analysis of present day American English. Providence, Rhode Island: Brown University Press

Lorge, I. and Chall, J. 1963: Estimating the size of vocabularies of children and adults: an analysis of methodological issues. Journal of Experimental Education 32, 147-157

Lynn, R. W. 1973: Preparing word lists: a suggested method. RELC Journal 4, 25-32

McKeown, M. G., Beck, I. L., Omanson, R. C. and Pople, M. T. 1985: Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. Reading Research Quarterly 20, 522-535

Melka Teichroew, F. J. 1982: Receptive versus productive vocabulary: a survey. Interlanguage Studies Bulletin 6, 5-33

Nagy, W. E. and Anderson, R. C. 1984: The number of words in printed school English. Reading Research Quarterly 19, 304-330

_____, Herman, P. A. and Anderson, R. C. 1985: Learning words from context. Reading Research Quarterly 20, 233-253

Nation, I. S. P. 1983: Testing and teaching vocabulary. Guidelines 5, 12-25

_____ 1984a: Teaching and learning vocabulary. Revised edition. Wellington, New Zealand: English Language Institute, Victoria University

Nation, I. S. P. 1984b: Vocabulary lists: words, affixes and stems. Wellington, New Zealand: English Language Institute, Victoria University

Oskarsson, M. 1980: Approaches to self-assessment in foreign language learning. Oxford: Pergamon

Praninskas, J. 1972: American university word list. London: Longman

Quinn, G. 1968: The English vocabulary of some Indonesian university entrants. Salatiga, Indonesia: IKIP Kristen Satya Watjana

Richards, J. C. 1976: The role of vocabulary teaching. TESOL Quarterly 10, 77-89

Seashore, R. H. 1933: The measurement and analysis of extent of vocabulary. Psychological Bulletin 30, 709-710

Sims, V. M. 1929: The reliability and validity of four types of vocabulary tests. Journal of Educational Research 20, 91-96

Tilley, H. C. 1936: A technique for determining the relative difficulty of word meanings among elementary school children. Journal of Experimental Education 5, 61-64

von Elek, T. 1982: Test of Swedish as a second language: an experiment in self-assessment. Work Paper No.31. Göteborg, Sweden: Language Teaching Research Center

West, M. 1953: A general service list of English words. London: Longmans, Green

Xue, Guo-yi and Nation, P. 1984: A university word list. Language Learning and Communication 3, 215-229

Zettersten, A. 1985: Experiments in large-scale vocabulary testing. In Lee, Y. P., Fok, A. C. Y. Y., Lord, R. and Low, G., editors, New directions in language testing. Oxford: Pergamon